DOCUMENT RESUME

ED 443 319 HE 033 074

AUTHOR Huynh, Cam-Loi; Hladkyj, Steve
TITLE Opinions of Administrators and Faculty on the Purposes, Control and Process of Performance Indicators in Higher Education: A Pilot Study.
PUB DATE 2000-04-00
NOTE 58p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Accountability; *Administrator Attitudes; College Faculty; Educational Quality; Educational Researchers; Evaluation Methods; Foreign Countries; Higher Education; *Performance Based Assessment; Tables (Data); *Teacher Attitudes; *Teacher Evaluation
IDENTIFIERS *Canada; *Performance Indicators

ABSTRACT

This study investigated the opinions of college faculty and administrators regarding the purpose, control, and process of performance evaluation, hypothesizing that job orientations and expectations would influence their opinions--that administrators would favor an economic model emphasizing authoritative and quantitative measures; teachers would favor an information model emphasizing networking relationships; and researchers would favor a hybrid approach. Questionnaires sent to three Canadian universities and completed by administrators, professors, instructors, and researchers yielded 116 usable replies. The questionnaire's 54 items focused on: purposes of performance evaluation; control and process of performance evaluation; standards; validity of performance indicators; overall opinions on the issues of purposes, control, and process, as well as satisfaction with existing performance evaluation systems; and demographics. Results indicated that job orientation and expectations of respondents influenced their views on purposes, control sources, and implementation procedures of performance evaluation. Administrators favored an economic model; teachers favored an information model; and researchers favored a hybrid approach. Respondents believed evaluation should be annual for nontenured faculty and every two to three years for tenured faculty. There was substantial agreement among respondents about the appropriate list of performance indicators. Data tables and diagrams of rank-ordered means are appended. (Contains 19 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 443 319

Opinions of Administrators and Faculty on the Purposes, Control and Process of Performance Indicators in Higher Education: A Pilot Study[1]

Cam-Loi Huynh[2] and Steve Hladkyj

Department of Psychology

University of Manitoba

Winnipeg, Manitoba Canada R3T 2N2

Key Words: education indicators, education policy, economic model, information model

[1] Paper presented at the annual meeting of the American Educational Research Association, April 24-28, 2000, New Orleans, LA.

[2] The authors can be contacted at Huynh@cc.umanitoba.ca

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

C-L Huynh

BEST COPY AVAILABLE

Abstract

This is an empirical investigation on the differential effects of the Economic and Information models in identifying purposes, control sources and implementation procedures of a performance evaluation system in higher education. The t test of means difference, one-way ANOVA F and structural groupings in factor analysis reveal that, due to their varied job orientations and expectations, administrators tend to associate with the characteristics of the Economic model, teachers with the Information model and researchers, with a mixture of both. Nevertheless, there was a substantial agreement among those surveyed about the appropriate list of performance indicators based on the ranking of sample means.

Opinions of Administrators and Faculty on the Purposes, Control and Process of Performance Indicators in Higher Education: A Pilot Study

In recent years, there has been increasing concern and pressure arising from all levels of society for quality assessment and public accountability in higher education. Central to the quality assessment of higher education is the performance evaluation of faculty. Academic institutions (universities, faculties, departments, etc.) may have diverse values reflecting differing conceptions of higher education and the knowledge of these values is essential in determining the form and function of performance evaluation procedures. Naturally, different faculties and departments within any given institution may employ various forms of evaluation. The conflict in these dimensions has been demonstrated in the European experience over the last decade (Cave, Hanney, Henkel, & Kogan, 1997). In this study, we introduce two other research aspects, namely, the differential views on demand characteristics of the academic activities held by administrators and faculty; and, different job orientations and expectations (research, teaching, service, etc.) among the faculty, even in the same faculties or departments. These differences may require varied purposes and forms of evaluation. No empirical evidence on these issues has been reported in the North American context although it is well known that goals and procedures in faculty evaluation are diverse. Therefore, performance evaluation may become an important source of apprehension and misconception in academic settings (Newson, 1995). It is the primary objective of this study to conduct a survey of opinions held by faculty and administrators in a sample of three universities concerning the purpose, control and process of performance evaluation.

Theoretical Framework and Hypotheses

Performance Indicators (PIs) are defined by Darling-Hammond (1992) as "individual or composite statistics that reflect important features of a system, such as education, health, or the economy" (p. 236). More specifically, PIs are often defined as quantitative measures of some important aspects of university operations relative to institutional goals (Borden & Bottrill, 1994; Cave, Hanney & Kagan, 1991; Kells, 1990). Unfortunately, as Benjamin (1996) pointed out, this definition is problematic in at least four aspects: (i) it mistakenly implies that quantitative measures are better than qualitative ones, (ii) it may ignore the different needs of administrators and faculty in the implementation of performance evaluation, (iii) it tends to sacrifice the specific and itemized nature of PIs in protecting the generality of institutional goals, and (iv) it could erroneously treat all PIs as equally useful. In short, the definition and construction of PIs in any educational setting should be more empirically oriented than theoretically driven. This is the approach we follow here. In particular, we will empirically examine the relevance of two models in developing the performance indicators (PIs) in higher eduction: the Economic model (proposed by Cave and his colleagues) and the Information, or Communicative Action, model (proposed by Habermas, 1989, 1991 and reintroduced by Barnett, 1994).

In the Economic model, the conception of performance indicators (PIs) derives from the fact that the educational system can be considered as a process within a wider economic system which converts inputs (such as time spent on duties, faculty salaries, etc.) into output (such as number of graduates, research publication, etc.). In this model, PIs in higher education may be described as authoritative and quantitative measures of attributes of the

activities of institutions and component units (Cave et al., 1991). They entail the collection of data at different levels of aggregation to aid forming judgments on faculty performance--judgments which may be made either within departments, within institutions or at the level of higher education system as a whole. Concrete steps in the Economic model can be specified either by the Production approach (Nedwek & Neal, 1994) or the Student Development approach (Pascarella & Terenzini, 1991). In the Production system, universities are considered as manufacturing plants that transform entering students into graduates. Variants of the production approach are the only ones in current use in Europe and North America (Benjamin, 1996; Borden & Bottrill, 1994; Gaither, Nedwek & Neal, 1994; Nedwek & Neal, 1994). Under the Student Development approach, PIs are designed to facilitate the attainment of "whole students" who not only know course and program contents (cognitive indicators) but are also integrated into academic and social communities (maturation indicators). In short, all variants of the Economic model would lead to an outcome-based performance evaluation system.

In the Information model, performance evaluation in higher education is quality-oriented, serving different ends ("Purpose"), being conducted by separate parties ("Control") and employing different techniques ("Process"). It emphasizes a networking relationship that takes into consideration the needs, as well as the contribution, of all participants in the system. Barnett (1994) explained this model in terms of three conceptual axes: enlightenment, power and form. For an example, see Figure 1. First, the *purpose* dimension spans along the enlightenment axis. At one end, quality improvement is embedded in the premise that understanding will be maximized where it is self-oriented ("Emancipatory"). At the other end,

quality evaluation is aimed at external auditing and recognition ("Technicist"). Secondly, the *control* dimension is represented along the power axis. At its opposite ends are methodologies essentially under the control of either the professional staff ("Collegial") or authorities external to them ("Administrative"). The third analytical dimension is the *process* of evaluation which can be analyzed along the form axis. Its two extremes consist of methods for quality assessment against the practical or gold standard of performance ("Bureaucratic") versus methods of quality improvement based on a universal or logical ideal ("Professional"). Essentially, in comparing the two models, it seems that the Economic model is mainly concerned with the endpoints along the continuums conceptualized in the Information model. Items reflecting both of these models were included in the questionnaire survey.

Insert Figure 1 about here

Ramsden (1991) has argued that questions on PIs taken out of context or data analysis based on individual responses are often misleading. Moreover, there are difficulties in interpreting students' evaluations of individual faculty members (Pollitt, 1990; Warnock, 1989). Therefore, in this study, the questions are addressed directly to the faculty and administrators which included items related their opinions about the performance evaluation system as a whole. Because PIs are essentially about the relative performance of aggregates, it is necessary that the question items and results are designed and analyzed at the aggregate level.

In psychological experiments, demand characteristics refer to the cues and other

information used by the participants to guide their behaviors, or responses (Orne, 1962). In our questionnaire survey, it is explained that, on the basis of the demand characteristics of academic activities, academic workers in higher education can be divided into three groups: Administrators, Teachers and Researchers. Respondents were asked for self classification into one of these job categories. Although most respondents may engage in all of these activities, the involvement is often at various degrees. Therefore, it was hypothesized that the emphasis on their own job orientations and expectations would influence their opinions on the purposes, control and procedures of a performance evaluation system. It was also hypothesized that Administrators would favor the Economic model, Teachers would support the Information model and Researchers would prefer a hybrid approach.

Methods

Questionnaire

The questionnaire consists of 54 items grouped into six parts: (i) purposes of performance evaluation (16 items: P1 to P16), (ii) control and process of performance evaluation (21 items: C17 to C36. Item C37 for open comments is deleted), (iii) standards (1 item: I38A to I38O, for 14 performance indicators and an open option), (iv) validity of PIs (2 items: I39A to I39N and I40A to I40N, for "objective" and "subjective" validity of the 14 PIs, respectively), (v) overall opinions on the issues of purposes (P41), control (C42) and process (I43) as well as satisfaction (S44) of existing performance evaluation systems, and (vi) demographic information (10 items: D1 to D10, for job orientation, professional backgrounds, age, gender, annual income, external grants and subordinates, etc.). With three exceptions, all

items in the first four parts are designed as Likert-typed with 7 scales (1 to 7 with 4 as the neutral anchor). Items C17 and C18 ask about frequency of performance (from 1 to 4 years plus "Other"). Item I38 ask respondents to weight each of the 14 PIs (plus "Other", if necessary) that must sum to 100%. The four items in part (v) represent the continuum with 10 anchor points to reflect the axes in Habermas (1989)'s Information model.

Data Collection

Three hundred questionnaire forms were sent to three universities in a prairie province of Canada: a large doctoral university with two campuses in the same city, one medium-sized urban non-doctoral university and one small rural four-year institution. Two populations in higher education were surveyed: "Administrators" (e.g., presidents, vice presidents, deans, associate deans, directors and department heads) and "Faculty" (e.g., professors, researchers, instructors) in Arts and Humanities. Addresses were recorded from the Internet Web sites of the universities, university catalogues and telephone directories. Questionnaire forms were mailed to all identified authorities for the "Administrators" sample. For the "Faculty" sample, questionnaire copies were sent to identified faculty as well as heads of departments/units for distribution. A downloadable version of the questionnaire was also made available on the Internet Web page of the first author. With the exception of the information about their positions and institutions, the identification of respondents were unknown, and all individual responses were kept confidential.

Results

There were 125 returns of which 9 were unusable due to wrong addresses or unidentified recipients, resulting in 116 completed, usable forms. The number of returns

among the three universities, 65 (56%), 32 (28%) and 19 (16%), respectively, resembles the relative distributions of the targeted populations. Seventy six percent of respondents were males, 89% were employed full-time and 39% had external grants. On the average, the typical respondent had 20 years of experience, between 45 to 55 years old, with an annual income between $60,000 to $70,000 (Canadian funds) and had about 4 to 5 students or employees under direct supervision. They spent 41% of time on teaching, 30% on research, 11% on department-level service, 6% on faculty-level administration and 12% on other activities (university-level, community service etc.). Responses in the total sample were sorted into three subsamples according to the self-reported job orientation: Administrators, Teachers and Researchers. There were 10 respondents who failed to identify their job orientation, resulting in 106 usable observations for the subsamples. Demographic characteristics of the total and subsamples are given in Table 1. The division in job orientations and expectations is reflected in how working time was spent in the three subgroups: 65% of Administrators' time was spent for service (Department-level, Institutional-level and Others), 50% of Teachers' time for Teaching and 53% of Researchers' time for Research.

Insert Table 1 about here

In the following, data analysis was conducted by means of descriptive statistics, $\underline{t}$ and ANOVA $\underline{F}$ tests, and factor analysis. Main results, and explanations of variables when necessary, are given in the tables. In an effort to find the most relatively important variables in each table, their arithmetic averages were ranked and compared within the total sample and

subsamples. Then, two statistical procedures were used for evaluating the hypothesis that Administrators favor the Economic model whereas Teachers, and to a lesser degree, Researchers support the Information model. First, one-way ANOVA F was analyzed to detect effects of job orientations and expectations across independent subsamples. Subsequently, latent classes formed across samples by factor analysis with varimax rotation were interpreted in light of their grouped variable components. For this purpose, the factorial groups were labelled on the basis of their component constitution. All statistical tests were evaluated at α = .05 (with Bonferonni adjustment for t test statistics).

Relative Importance of PI Purposes. Nine potential goals of a performance evaluation system are listed in Table 2. Each of these items has seven scales, the first three ratings (1, 2 and 3) indicate "Unimportant" and the last three (5, 6, 7) signify "Important". From the total sample, they are grouped into three latent classes: "Improvement/Emancipatory" (Teaching, Research, Service), "Development" (Promotion, New and Tenured Faculty), and "Comparison/Technicist" (Intra-department, Inter-department and Inter-university).

From the ordered item means, the components of "Comparison" are relatively unimportant (means < 4) whereas the three most important goal variables (means > 5) are identified as Teaching, New Faculty and Promotion. Only the rating of Service is significantly different among the three subsamples of Administrators (largest, mean > 5.3), Teachers and Researchers (smallest, mean < 3.5) according to the omnibus ANOVA F test.

Results from the Economic model (presumably represented by Administrators' responses) were quite different from those of the Information model (presumably supported by Teachers). First, the variables of improvement and monitoring in Research, Service and

Tenured Faculty were rated highest in the Administrators sample (means > 5). Secondly, whereas the latent grouping of the Teachers sample was the same as that of the total sample, the Administrators sample yields five classes: "Comparison-plus" (the three "Comparison" components plus Teaching), "Evaluation" (Research, New Faculty) and the remaining three latent classes, each has only one component. Researchers present a mixture of these two models, with Intra-department moves out of "Comparison" and into "Development." Moreover, Research was not in "Development" but became a class by itself. Ordered means and factorial groupings are depicted along the three axes of Administrators, Teachers and Researchers in Figure 2.a.

Insert Table 2 about here

Institutions that Determine PI Purposes. Participants were asked to rate the relative leverage of seven institutional bodies in making decision on purposes of performance evaluation (Table 3). From all samples, Department-level unit was consistently rated highest (mean > 5.0) whereas public institutions (Government and Consulting Agency) were rated smallest (means < 2.0) (Figure 2.b). The rating of University-level is significantly different among the three subsamples according to the omnibus F test (rated highest by Administrators and lowest by Researchers).

The seven institutions in Table 3 were grouped into four latent classes in the total sample: "Union/Emancipatory" (Faculty Union), "Authority/Technicist" (Board of Governors, University-level), "Professional" (Faculty-level, Department-level), and "Public" (Government

and Consulting Agency). The disparity in the implications of the Economic and Information models can be detected by the fact that only Administrators supported Faculty-level determination of PI purposes; and the grouping of latent classes for the Administrators and Teachers samples are different (Figure 2.b). Although all other groupings in the total sample were maintained, Department-level was reclassified from "Professional" to "Union" in the Administrators sample, implying that departmental units would collaborate with Faculty Union (Emancipatory) whereas all other identities represented different levels of authority (Technicist). The four classes from the Teachers sample were: "Union/Emancipatory" (Union), "Higher Authority/Technicist" (Government, Board of Governors), "Lower Authority/Technicist" (University-level, Faculty-level) and "Professional/Emancipatory" (Department-level, Consulting Agency). Researchers reproduced two classes of Teachers' grouping and modified the other two as "Authority 1" (Board of Governors, Faculty-level) and "Authority 2" (University-level, Department-level, Government).

Insert Table 3 about here

Frequency of Conduct the Performance Evaluation. All samples indicated that the evaluation frequency should be annually for non-tenured faculty and from two to three years for tenured faculty (Table 4).

Insert Table 4 about here

"Triggers" of Performance Evaluation. From a list of 11 possible "triggers" (Table 5) by which a formal evaluation process could be initiated, only two were supported (means > 5), Self-request or Automatically (i.e., by the passage of a calendar time). However, Researchers rated only Automatically above 5 among all triggers (Figure 2.c).

From the total sample, 5 latent classes were found: "Routine" (Automatically), "Concerned Agents/Collegial" (Self Request, Peers, Students, Union), "Lower Authority/Administrative" (Department Head, Department-level, Faculty-level), "Higher Authority/Administrative" (University-level, Government) and "Public" (Granting Agencies). Teachers reproduced this grouping except that Students was found in "Public." On the other hand, there were eight classes in the Administrators sample. Besides "Lower Authority/Administrative" (without Faculty-level), and "High Authority/Administrative", the rest were single-component classes (Figure 2.c). Researchers conceptualized only four latent groups: "Routine" (Automatically), "Concerned Agents 1" (Self Request, Union, Government, Granting Agencies), "Concerned Agents 2" (Peers, Students) and "Authority/Administrative" (Department Head, Department-level, University-level, Faculty-level).

Insert Table 5 about here

Administrators of Performance Evaluation. There are eight institutions that presumably can administer the routine process of performance evaluation (Table 6). Results from all samples imply that the implementation of performance evaluation should be a Department-level responsibility. As expected, Faculty-level was also supported by Administrators (means > 5)

(Figure 2d). The rating of Faculty Union was significantly different among the three subsamples according to omnibus F test (with smallest rating by Administrators and largest by Researchers). Since the rating of Teachers is much closer to that of Researchers, this implies the underlying disparity of Economic and Information models on this issue.

The four latent classes in the total sample were identified as "Union/Collegial" (Union), "Public" (Licencing Body, Government, Consulting Agency), "Higher Authority/Administrative" (Board of Governors, University-level), and "Lower Authority/Administrative" (Department-level, Faculty-level). In the Administrators sample, Licencing Body was assigned to "Higher Authority." Besides these two groups, the remaining variables form one-component classes. Teachers reproduced the classification of the total sample, except that "Lower Authority" disappeared since Government and Department-level formed a class and Faculty-level was moved into "Higher Authority." The latent grouping in the Researchers sample is quite similar to that of Teachers (Figure 2.d).

Insert Table 6 about here

Holders of Individual Performance Information. If a formal performance evaluation had been implemented, which institutions or identities could get access to the final information besides the individual faculty involved? Table 7 lists nine potential recipients. Access by university central administration (University-level) was endorsed in three samples (mean > 5), but clearly opposed by Teachers (mean = 1). On the other hand, Faculty Union access was supported by both Administrators and Teachers (means > 5) but only weakly by Researchers

(mean < 4.5) (Figure 2e).

The nine variables were grouped into eight classes by factor analysis in three samples (except Administrators), with only one multiple class: "Concerned Agents" (Students, Private Agencies). In the Administrators sample, six latent groups were formed, with two multiple factors: "Concerned Private" (Student, Department Head, Private Agencies) and "Concerned Public" (Public, Government). This again implies a structural difference in the implications of the Economic (Administrators) and Information (Teachers and Researchers) models (Figure 2.e).

Insert Table 7 about here

The Fourteen Performance Indicators: Relative Weights of Importance. Respondents were asked to assign a percentage to each of 14 PIs (summed to 100%) according to its relative importance in a performance evaluation system (Table 8). The two PIs uniformly rated with weights more than 10% were Course Evaluation and Book Publication. Added to this list were Grants in both Administrators and Researchers subsamples; and Peer-reviewed Journal Publication in Teachers and Researchers subsamples (Figure 2f). The three variables of Grants, Peer-reviewed Journal Publication and Number of Courses Taught were significantly different among the three subsamples according to omnibus F test. The ranking of ordered means can be used to explain this outcome as well as to analyze the departure of Administrators' opinions from the those of Teachers and Researchers. Grants were weighted much larger by Administrators and Researchers than by Teachers. Number of Courses Taught

was weighted highest by Administrators and lowest by Researchers whereas Peered-reviewed Journal Publication was rated much higher by Teachers and Researchers than by Administrators.

Both the total and Teachers samples yield eight latent classes with three of them having multiple variables. In the total sample, they are "Formal Achievement" (Grants, Peer-reviewed Conference Presentation, Peer-reviewed Journal Publication, Book Publication), "Informal Achievement" (Non-peer reviewed Conference Presentation, Non-peer reviewed Journal Publication) and "Teaching Quality" (Student Supervision, Graduate Success). Besides "Formal Achievement", the two multiple-component classes formed by Teachers are "Recognition 1" (Non-peer reviewed Conference Presentation, Reputation) and "Recognition 2" (Non-peer reviewed Conference Presentation, Award). The Researchers sample yields nine classes but the groupings are still quite similar to those in the Teachers sample (Figure 2.f). Finally, there are 10 latent classes in the Administrators sample with three multiple groups: "Recognition" (Years of Experience, Reputation), "Formal Achievement" (Grants, Peer-reviewed Journal Publication, Book Publication), and "Informal Achievement" (as above).

Insert Table 8 about here

The Fourteen Performance Indicators: "Subjective" Validity. Respondents were asked to rate the relative suitability of 14 PIs to their individual and institutional situations (Table 9). Peer-reviewed Journal and Book publications were rated highest (means > 5), and Graduate Success was rated lowest (mean < 3), with respect to "subjective validity" in all samples.

Added to this list were Grants (rated by Administrators; mean > 5); and Years of Experience and Non-peer reviewed Conference Presentation (rated by Researchers; means < 3) (Figure 2.g). The omnibus F tests were significant for Peer-reviewed Journal Publication (largest by Researchers and smallest by Administrafors) and Number of Courses Taught (largest by Teachers and smallest by Researchers).

Seven latent classes were found in the total sample, with four multiple classes, namely, "Formal Achievement" (Grants, Peer-reviewed Journal Publication, Book Publication), "Student Contribution" (Number of Students Supervised, Number of Courses Taught), "Work Quality" (Community Service, Course Evaluation by Students) and "Recognition" (Graduate Success, Reputation, Award). Besides "Formal Achievement" and "Student Contribution", the other two multiple classes, out of a total of six, in the Teachers sample were "Informal Achievement" (Non-peer reviewed Conference Presentation, Non-peer reviewed Journal Publication) and "Contribution-plus" (Peer-reviewed Conference Presentation, Community Service, Course Evaluation, Award). Researchers conceptualized seven latent classes with four multiple groupings (Figure 2.g). Besides "Formal Achievement" and "Work Quality" as above, the other groups were "Informal Contribution" (Non-peer reviewed Conference, Students-supervised), and "Recognition" (Graduate Success, Reputation). On the other hand, there were 10 latent classes in the Administrators sample with only one multiple group of "Achievement-plus" (Grants, Peer-reviewed Journal Publication, Peer-reviewed Conference Presentation, Number of Students Supervised).

Insert Table 9 about here

The Fourteen Performance Indicators: "Objective" Validity. Respondents were asked to rate the relative suitability of the listed PIs to general academic settings (Table 10). The PIs that received highest ratings (means > 5) for "objective" validity in all samples were Peer-reviewed Journal Publication, Peer-reviewed Conference Presentation, Book Publication, Number of Students Supervised, and Reputation. As expected, Grants were highly rated by Administrators (Figure 2h). Both Grants (largest for Administrators) and Number of Courses Taught (largest for Teachers and smallest for Researchers) were statistically significant by the omnibus F test. On the other hand, Award, Course Evaluation, and Graduate Success were unanimously dismissed as objectively-valid performance indicators across all samples (means = 1).

Seven latent classes were formed in the total sample, with four multiple groups, namely, "Formal Achievement" (Grants, Peer-reviewed Journal Publication, Peer-reviewed Conference), "Informal Achievement" (Years of Experience, Non-peer reviewed Conference Presentation, Non-peer reviewed Journal Publication), "Teaching Quality" (Graduate Success, Course Evaluation by Students), and "Recognition" (Number of Courses Taught, Reputation). Eight classes were factorially grouped in the Administrators subsample, with three multiple groups of "Achievement-plus" (Grants, Peer-reviewed Journal Publication, Peer-reviewed Conference Presentation, Number of Courses Taught, Reputation), "Informal Achievement" (Non-peer reviewed Conference Presentation, Non-peer reviewed Journal Publication), and "Teaching Quality" (as above). The Teachers sample produced 13 classes with only one multiple group (namely "NPR Achievement"). Among the 12 factorial classes in the Researchers sample, two were multiple groups, namely "Informal Achievement" and "Course

Contribution" (Number of Courses Taught, Course Evaluation by Students) (Figure 2.f).

Insert Table 10 and Figure 2 about here

It is of interest to study the difference in the two sets of validity ratings. The t test statistics of means difference for independent groups were found significant for the ratings of Number of Students Supervised, Graduate Success, Course Evaluation by Students and Reputation across all samples. This statistical significance implies that the roles and validity of these PIs for performance evaluation varied relative to the identified parties, namely, whether they were viewed by individuals or institutions. Only the variable of Non-peer reviewed Journal Publication is significantly different among the three subsamples according to the omnibus F test on difference scores. However, this finding is not meaningful due to its low ratings in all samples (mean < 4 for both subjective and objective validity settings).

Insert Table 11 about here

Structural Factors of A Performance Evaluation System. What are the overall characteristics of a system of performance evaluation in higher education if they are measured along the three axes of Purposes, Control, Procedures as conceptualized by Habermas (1989)? (Table 12). For the scales from 1 to 10, respondents seemed to feel neutral between "Evaluative/Technicist" and "Informational/Emancipatory" along the Purposes continuum (5 < means < 6), showed a slight preference for "Internal/Collegial" control rather than

"External/Administrative" control (means < 4) and were indifferent in the choice between "Professional" and "Institutional/Bureaucratic" procedures or standards of evaluation (means = 5). These findings are found uniformly across all samples. Is the existing system of performance evaluation satisfactory? Teachers and Researchers tended to be neutral (means $\geq$ 5) whereas Administrators seem to be somewhat dissatisfied (mean < 5).

Insert Table 12 about here

Summary and Conclusions

From the survey findings, it is evident that a performance evaluation system should be designed with the aim of improving teaching, monitoring the development of new faculty and making promotion and tenure decisions. It should not be used for comparison at departmental or higher levels. The goals of a performance evaluation system should be determined at departmental level and independent of governmental or private third-party influence. It is sufficient to conduct an evaluation annually for non-tenured faculty and at least every two-year interval for tenured faculty. Besides routine and informal performance review, it is only appropriate to open a formal evaluation upon a self request by the faculty. The administration of such a formal evaluation is most suitable if it is administered at the departmental level. It is preferred that only the individuals themselves could have access to information about their own performance evaluation.

There is a widely-held belief that performance quality in higher education is a many-

sided yet ultimately elusive phenomenon. This conviction has led several researchers to doubt whether unambiguous scales of measurement suitable as PIs could ever be derived (Cave et al., 1988; Smith, 1988). This conclusion seems altogether too pessimistic. Although how to identify "good" performance indicators and performance evaluation systems is undoubtedly a complicated matter, this study shows a substantial measure of agreement among those surveyed about their essential characteristics. The most relevant PIs were found to be publication in peered-reviewed journals and books as well as course evaluation by students. Although typical respondents considered publication records (peer-reviewed journals and books) as the most appropriate performance measures for themselves, other indicators such as peer-reviewed conference participation, number of students supervised, and reputation among peers would also be suitable in general. On the other hands, such PIs as award, merits, public recognition, and career success of former students were unanimously dismissed as objectively-valid performance yardsticks. The role of course evaluation by students as a performance indicator is unclear. It was rated highest for its relative importance among 14 PIs in all samples except Researchers, and lowest as a subjectively-valid measure of performance in all samples.

This study demonstrates that job orientations and expectations of participants in higher education would influence their views on purposes, control sources and implementation procedures of a performance evaluation system. It is evident that Administrators would favor the Economic model, Teachers would support the Information model and Researchers would prefer a hybrid approach of both.

In this pilot study, our conclusions may only apply to the three universities involved.

However, it is hoped that our findings, based on relatively long and sophisticated questionnaire, will provide some answers to Benjamin (1996)'s four issues with respect to the commonly-held definition for PIs presented previously. This model-based study will set the stage for our next investigation in which the same questionnaire will be sent to a larger number of departments and universities.

References

Barnett, R. (1994). Power, enlightenment and quality evaluation. European Journal of Education,29, 165-179.

Borden, V. M. H. & Bottrill, K. V. (1994). Performance indicators: History, definitions, and methods. In V.M. H. Borden, & T. W.Banta (eds.). Using performance indicators to guide strategic decision making (New Directions for Institutional Research, No. 82, pp. 5-21). San Francisco: Jossey-Bass.

Cave, M. and Hanney, S. (1990). Performance indicators for higher education and research. In M. Cave, M. Kogan and R. Smith. Output and performance measurement in government: The state of the art. London, UK: Jessica Kingsley Publishers.

Cave, M. and Hanney, S. (1995). Performance measurement in higher education and research-revisited. Public Money and Management, 15(4), 17-23.

Cave, M., Hanney, S., Kogan, M. and Trevett, G. (1991). The use of performance indicators in higher education: A critical analysis of developing practice (2nd ed.). London, UK: Jessica Kingsley Publishers.

Cave, M., Hanney, S., Henkel, M. and Kogan, M. (1997). The use of performance indicators in higher education. The challenge of the quality movement (Third edition). London, UK: Jessica Kingsley Publishers.

Darling-Hammond (1992). Educational indicators and enlightened policy. Educational Policy, 6, 235-265.

Kells, H. R. (ed.) (1990). The development of performance indicators for higher education (2nd ed.). Paris: Organization for Economic Co-operation and Development.

Gaither, G., Nedwek, B. P. & Neal J. E. (1994). Measuring up: The promises and pitfalls of performance indicators in higher education. ASHE-ERIC Higher Education Report. No. 5. Washington, DC: The George Washington University, Graduate School of Education and Development.

Habermas, J. (1989). The Theory of Communicative Action. Cambridge, UK: Polity.

Habermas, J. (1991). The Theory of Communicative Action (Vol. 1) Cambridge, UK: Polity.

Nedwek, B. P. & Neal, J. E. (1994). Performance indicators and rational management tools: A comprehensive assessment of projects in North America and Europe. Research in Higher Education,35, 75-103.

Newson, J. A. (1995). Presidential address: Positioning the social sciences in a context of economic restructuring. Society-Societe, 19, 1-11.

Orne, M. T. (1952). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist, 17, 776-783.

Pascarella, E. T. & Terenzini, P. T. (1991). How college affects students: Findings and insights from 30 years of research. San Francisco: Jossey-Bass.

Pollitt, C. (1990). Measuring university performance: never mind the quality, never mind the width? Higher Education Quarterly, 44, 61-81.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: the course experience questionnaire. Studies in Higher Education, 16(2), 129-150.

Smith, D. M. (1988). On academic performance. Area, 20, 3-13.

Warnock, M. (1989). A common policy for education. Oxford, UK: Oxford University Press.

Table 1. Demographic characteristics of survey respondents[1]

Variable	Total N = 116 Mean (%)	Administrs N = 16 Mean (%)	Teachers N = 66 Mean (%)	Researchers N = 24 Mean (%)
D1. PLace of Employment				
Large University	65	8	36	14
Medium University	32	7	17	5
Small University	19	1	13	5
D2. Employment Status				
Full-time	103	16	60	19
Non Full-time	13	0	6	5
D3. Years Employed	19.90	15.71	21.14	20.52
D4. Age	54	50	55	53
D5. Gender				
Males	88	15	46	18
Females	28	1	20	6
D6. Annual Income	65K	60K	70K	60K
D7. External Grants	(0.39)	(0.19)	(0.44)	(0.32)
D8. Subordinates	4.68	2.31	3.52	10.13
D9. Time spent on duties				
Teaching	(0.41)	(0.23)	(0.50)	(0.28)
Research	(0.30)	(0.09)	(0.26)	(0.53)
Department-level	(0.11)	(0.21)	(0.10)	(0.09)
Faculty-level	(0.06)	(0.03)	(0.07)	(0.07)
Institutional-level	(0.05)	(0.13)	(0.03)	(0.02)
Others	(0.07)	(0.31)	(0.04)	(0.01)

[1] In this and all subsequent tables, the sum of subsample frequency counts in each row is not equal to the total sample count due to 10 missing values for self-reported job orientation.

Table 2. Relative importance of the various purposes served by performance evaluation.

Variable		Total N = 116 Mean (SD) Factor	Administrs N = 16 Mean (SD) Factor	Teachers N = 66 Mean (SD) Factor	Researchers N = 24 Mean (SD) Factor	F (p)
P1	Teaching	5.71 F1 (1.38)	6.25 F1 (1.24)	5.66 F1 (1.45)	5.43 F1 (1.31)	1.70 (.189)
P2	Research	4.41 F1 (2.08)	5.37 F2 (1.93)	4.23 F1 (2.07)	4.46 F2 (1.93)	2.07 (.131)
P3	Service	3.97 F1 (1.96)	5.40 F3 (1.24)	3.91 F1 (1.96)	3.43 F1 (1.83)	5.46 (.006)
P4	Promotion	5.17 F2 (1.69)	5.19 F4 (1.64)	5.01 F2 (1.76)	5.37 F3 (1.66)	0.40 (.673)
P5	New Faculty	5.30 F2 (1.52)	5.31 F2 (1.35)	5.17 F2 (1.60)	5.37 F3 (1.55)	0.18 (.836)
P6	Tenured Faculty	4.88 F2 (1.73)	5.19 F5 (1.47)	4.83 F2 (1.80)	4.79 F3 (1.59)	0.31 (.731)
P7	Intra-department	3.27 F3 (1.89)	2.93 F1 (1.53)	3.18 F3 (1.78)	3.42 F3 (2.04)	0.34 (.714)
P8	Inter-department	2.50 F3 (1.58)	2.80 F1 (1.78)	2.48 F3 (1.51)	2.37 F4 (1.34)	0.38 (.686)
P9	Inter-university	2.261 F3 (1.60)	2.40 F1 (1.76)	2.17 F3 (1.55)	2.50 F4 (1.64)	0.43 (.651)

P1 = to improve teaching. P2 = to improve research. P3 = to improve service. P4 = to make promotion & tenure decisions. P5 = to monitor development of new faculty overtime. P6 = to monitor continued performance of tenured faculty. P7 = to compare performance of individuals within departments. P8 = to compare performance between departments. and P9 = to compare performance quality between universities.

F = one-way omnibus test for ANOVA design of three subsamples. = p-value of F.

Table 3. Relative importance of institutions that determine purposes of performance evaluation.

Variable		Total N = 116 Mean (SD)	Factorial Grouping	Administrs N = 16 Mean (SD)	Factorial Grouping	Teachers N = 66 Mean (SD)	Factorial Grouping	Researchers N = 24 Mean (SD)	Factorial Grouping	F (p)
P10	Faculty Union	4.20 (2.28)	F1	3.33 (1.84)	F1	4.54 (2.27)	F1	4.23 (2.37)	F1	1.81 (.170)
P11	Board of Governors	2.47 (1.85)	F2	3.07 (2.19)	F2	2.39 (1.73)	F2	2.24 (1.67)	F2	1.06 (.349)
P12	University-level	3.12 (1.95)	F2	4.56 (2.06)	F2	2.89 (1.79)	F3	2.60 (1.82)	F3	6.21 (.003)
P13	Faculty-level	4.65 (1.78)	F3	5.37 (1.31)	F3	4.51 (1.85)	F3	4.45 (1.71)	F2	1.70 (.188)
P14	Department-level	5.19 (1.88)	F3	5.07 (1.94)	F1	5.18 (1.87)	F4	5.18 (1.89)	F3	0.02 (.976)
P15	Government	1.78 (1.39)	F4	1.73 (1.39)	F4	1.80 (1.44)	F2	1.67 (0.91)	F3	0.09 (.917)
P16	Consult Agency	1.85 (1.57)	F4	1.80 (1.78)	F4	1.94 (1.68)	F4	1.90 (1.37)	F4	0.04 (.957)

P10 = Faculty Union, P11 = Board of Governors, P12 = Central University Administration, P13 = Faculty-level Administration, P14 = Departmental-level Units/Committees, P15 = Government Department(s) of Education, P16 = Third-party Consulting Agency.

Table 4. Frequency of performance evaluation of non-tenured and tenured faculty

Variable		Total N = 116 Mean (SD)	Administrs N = 16 Mean (SD)	Teachers N = 66 Mean (SD)	Researchers N = 24 Mean (SD)	F (p)
C17	Non-tenured	1.32 (0.83)	1.31 (1.01)	1.37 (0.83)	1.26 (0.86)	0.16 (.855)
C18	Tenured	2.51 (1.27)	2.25 (1.39)	2.48 (1.23)	2.62 (1.31)	0.41 (.662)

C17 = Frequency to conduct a performance evaluation for non-tenured faculty. C18 = Frequency to conduct a performance evaluation for tenured faculty.

Table 5. Triggers of formal performance evaluation for informational purposes

Variable	Total N = 116 Mean (SD) Factorial Grouping	Administrs N = 16 Mean (SD) Factorial Grouping	Teachers N = 66 Mean (SD) Factorial Grouping	Researchers N = 24 Mean (SD) Factorial Grouping	F (p)
C19A Automatically	5.54 F1 (1.77)	5.62 F1 (2.00)	5.77 F1 (1.56)	5.097 F1 (1.88)	1.33 (.268)
C19B Self-request	5.56 F2 (1.74)	5.69 F2 (1.70)	5.78 F2 (1.63)	4.87 F2 (2.05)	2.37 (.099)
C19C Dept. Head	2.93 F3 (1.89)	3.12 F3 (1.59)	2.94 F3 (1.91)	2.78 F3 (1.95)	0.16 (.854)
C19D Department-level	3.15 F3 (1.88)	2.688 F3 (1.49)	3.540 F3 (1.94)	1.820 F3 (2.70)	2.52 (.086)
C19E University-level	2.159 F4 (1.66)	2.313 F4 (1.85)	2.000 F4 (1.42)	2.348 F3 (1.77)	0.54 (.582)
C19F Faculty-level	2.99 F3 (1.86)	3.00 F5 (1.63)	3.09 F3 (1.81)	2.43 F3 (1.80)	1.16 (.316)
C19G Peers	3.33 F2 (1.98)	3.19 F6 (1.87)	3.44 F2 (1.93)	3.087 F4 (2.17)	0.31 (.736)
C19H Students	4.07 F2 (2.20)	4.37 F7 (2.33)	4.16 F5 (2.17)	3.35 F4 (2.35)	1.36 (.261)
C19I Faculty Union	3.009 F2 (2.05)	2.750 F7 (1.88)	2.953 F2 (194)	3.304 F2 (2.32)	0.40 (.671)
C19J Government	1.442 F4 (0.92)	1.438 F4 (1.26)	1.484 F4 (0.89)	1.478 F2 (0.95)	0.02 (.985)
C19K Granting Agencies	2.29 F5 (1.70)	2.75 F8 (2.41)	2.27 F5 (1.55)	2.13 F2 (1.66)	0.67 (.516)

C19A = automatically each year, C19B = on request by the faculty member, C19C = by department head, C19D = by a departmental committee, C19E = from a university-level administrative unit, C19F = from the faculty dean/director,C19G = by formal request from peers, C19H = by formal request from students, C19I = by faculty union, C19J = by government, C19K = by request from granting agencies.

Table 6. Ratings of institutions which can administer the routine process of performance evaluation

Variable		Total N = 116 Mean (SD) Factorial Grouping	Administrs N = 16 Mean (SD) Factorial Grouping	Teachers N = 66 Mean (SD) Factorial Grouping	Researchers N = 24 Mean (SD) Factorial Grouping	F (p)
C20	Faculty Union	3.27 F1 (2.29)	1.87 F1 (1.51)	3.52 F1 (2.28)	4.00 F1 (2.43)	4.51 (.013)
C21	Licencing Body	3.35 F2 (2.15)	2.67 F2 (1.88)	3.55 F2 (2.12)	2.83 F2 (2.15)	1.69 (.190)
C22	Board of Governors	2.09 F3 (1.66)	2.33 F2 (1.88)	2.09 F3 (1.68)	1.74 F3 (1.18)	0.68 (.507)
C23	University-level	2.58 F3 (1.97)	3.47 F2 (2.39)	2.44 F3 (1.87)	2.26 F3 (1.74)	2.07 (.131)
C24	Faculty-level	4.62 F4 (2.02)	5.53 F3 (1.64)	4.59 F3 (2.14)	4.43 F3 (1.70)	1.62 (.204)
C25	Department-level	5.22 F4 (1.74)	5.00 F4 (2.07)	5.28 F4 (1.28)	5.08 F2 (1.72)	0.22 (.801)
C26	Local Government	1.29 F2 (1.30)	1.47 F5 (0.68)	1.28 F4 (0.57)	1.35 F3 (.699)	(0.74)
C27	Consult Agency	1.78 F2 (1.65)	1.47 F5 (1.55)	1.96 F2 (1.87)	1.61 F4 (1.23)	0.74 (.482)

Table 7. Ratings of institutions which can access individual performance evaluation information

Variable		Total N = 116 Mean (SD)		Administrs N = 16 Mean (SD)		Teachers N = 66 Mean (SD)		Researchers N = 24 Mean (SD)		F (p)
			Factorial Grouping		Factorial Grouping		Factorial Grouping		Factorial Grouping	
C28	Students	3.64 (2.46)	F1	3.31 (2.65)	F1	3.95 (2.48)	F1	3.25 (2.29)	F1	0.93 (.399)
C29	Other Faculty	3.35 (2.15)	F2	2.67 (1.88)	F2	3.55 (2.12)	F2	2.83 (2.15)	F2	1.69 (.190)
C30	Department Head	2.09 (1.66)	F3	2.33 (1.88)	F1	2.09 (1.68)	F3	1.74 (1.18)	F3	0.68 (.507)
C31	Faculty-level	2.58 (1.97)	F4	3.47 (2.39)	F3	4.59 (2.14)	F4	2.26 (1.74)	F4	2.07 (.131)
C32	Faculty Union	4.62 (2.02)	F5	5.53 (1.64)	F4	5.28 (1.67)	F5	4.43 (1.70)	F5	1.62 (.204)
C33	University-level	5.22 (1.74)	F6	5.00 (2.07)	F5	1.28 (0.68)	F6	5.08 (1.72)	F6	0.22 (.801)
C34	Public	1.29 (0.74)	F7	1.47 (1.30)	F6	1.97 (1.87)	F7	1.35 (0.57)	F7	0.36 (.699)
C35	Government	1.78 (1.65)	F8	1.47 (1.55)	F6	1.97 (1.87)	F8	1.61 (1.23)	F8	0.74 (.482)
C36	Private Agencies	3.64 (2.46)	F1	3.31 (2.65)	F1	3.95 (2.48)	F1	3.25 (2.29)	F1	0.93 (.399)

Table 8. Weights (in percentages) of 14 performance indicators

Variable	Total N = 116 Mean (SD)	Factorial Grouping	Administrs N = 16 Mean (SD)	Factorial Grouping	Teachers N = 66 Mean (SD)	Factorial Grouping	Researchers N = 24 Mean (SD)	Factorial Grouping	F (p)
138A Experience	7.38 (18.85)	F1	3.76 (4.91)	F1	7.80 (20.17)	F1	8.56 (22.60)	F1	0.30 (.743)
138B Grants	8.41 (5.84)	F2	11.12 (7.34)	F2	6.96 (5.50)	F2	10.65 (4.84)	F2	5.47 (.006)
138C PR Conference	6.74 (3.83)	F2	5.76 (3.85)	F3	6.96 (3.87)	F2	6.52 (4.11)	F2	0.56 (.575)
138D NPR Conference	3.51 (3.49)	F3	2.48 (2.50)	F4	4.12 (3.78)	F3	2.61 (3.22)	F3	2.31 (.105)
138E PR Journal Pub	14.30 (9.76)	F2	9.69 (8.17)	F2	13.51 (9.59)	F2	18.17 (7.85)	F3	4.13 (.019)
138F NPR Journal Pub	3.69 (3.34)	F3	2.84 (3.16)	F4	4.07 (3.52)	F4	3.13 (3.22)	F4	1.14 (.323)
138G Book Publication	11.86 (8.28)	F2	10.05 (8.96)	F2	11.00 (7.38)	F2	15.13 (10.04)	F4	2.48 (.091)
138H Community Service	6.36 (5.20)	F4	6.29 (5.41)	F5	7.27 (5.39)	F5	4.52 (4.46)	F5	2.35 (.100)
138I Students-supervised	5.71 (4.92)	F5	6.02 (5.69)	F6	5.53 (4.85)	F6	4.78 (3.84)	F6	0.33 (.716)
138J Graduate Success	2.30 (3.87)	F5	1.50 (3.03)	F7	2.34 (4.28)	F7	2.26 (3.28)	F7	0.26 (.768)
138K Courses-taught	6.26 (9.56)	F6	11.88 (18.82)	F8	6.02 (7.51)	F8	3.04 (4.19)	F8	3.89 (.024)
138L Course Evaluation	13.15 (11.23)	F7	12.17 (11.47)	F9	14.58 (11.64)	F5	12.17 (10.75)	F5	0.52 (.599)
138M Reputation	3.74 (5.14)	F8	3.93 (5.94)	F1	3.84 (5.48)	F3	3.26 (4.16)	F5	0.11 (.893)
138N Awards	6.04 (4.78)	F8	6.07 (4.01)	F10	6.21 (5.25)	F4	5.96 (4.33)	F9	0.02 (.977)

Table 8 (continued)

Tables 8, 9, 10 and 11 contain the same following variables: A = Years of experience, B = External research grants, C = Peer-reviewed conference presentations, D = Non peer-reviewed conference presentations, E = Peer-reviewed journal publication, F = Non peer-reviewed journal publication, G = Book publication, H = Community service, I = Number of students currently supervised, J = Career success of former students, K = number of courses taught, L = Course evaluation by students, M = Reputation of faculty among peers, N = Awards, merits, public recognition.

Table 9. "Subjective" validity ratings of 14 performance indicators

Variable	Total N = 116 Mean (SD)	Factorial Grouping	Administrs N = 16 Mean (SD)	Factorial Grouping	Teachers N = 66 Mean (SD)	Factorial Grouping	Researchers N = 24 Mean (SD)	Factorial Grouping	F (p)
139A Experience	3.34 (2.04)	F1	3.00 (2.31)	F1	3.61 (1.99)	F1	2.83 (1.97)	F1	1.50 (.229)
139B Grants	4.40 (1.81)	F2	5.17 (1.47)	F2	3.98 (1.88)	F2	4.83 (1.63)	F2	3.48 (.035)
139C PR Conference	4.71 (1.65)	F3	4.83 (1.80)	F2	4.73 (1.63)	F3	4.50 (1.47)	F3	0.24 (.790)
139D NPR Conference	3.23 (1.70)	F4	3.42 (1.00)	F3	3.33 (1.71)	F4	2.68 (1.64)	F4	1.42 (.247)
139E PR Journal	5.78 (1.38)	F2	5.50 (1.38)	F2	5.61 (1.51)	F2	6.17 (0.70)	F2	1.70 (.189)
139F NPR Journal	3.53 (1.65)	F4	3.92 (0.79)	F4	3.45 (1.62)	F4	3.48 (1.88)	F1	0.42 (.656)
139G Book Publication	5.59 (1.67)	F2	5.50 (1.78)	F5	5.45 (1.74)	F2	5.71 (1.46)	F2	0.20 (.817)
139H Community Service	4.16 (1.77)	F5	4.08 (1.73)	F4	4.37 (1.80)	F3	3.65 (1.69)	F5	1.48 (.233)
139I Students-supvised	3.89 (1.83)	F6	3.67 (1.92)	F2	4.03 (1.82)	F5	3.26 (1.74)	F4	1.57 (.213)
139J Graduate Success	2.87 (1.78)	F7	2.58 (1.62)	F6	2.75 (1.70)	F6	2.92 (2.00)	F6	0.15 (.858)
139K Courses-taught	3.74 (1.74)	F6	3.83 (1.75)	F7	3.95 (1.67)	F5	2.74 (1.54)	F7	4.63 (.012)
139L Course Evaluation	4.77 (1.85)	F5	4.67 (1.87)	F8	4.84 (1.94)	F3	4.61 (1.67)	F5	0.15 (.859)
139M Reputation	3.82 (1.92)	F7	3.67 (1.72)	F9	3.81 (1.93)	F6	3.58 (1.95)	F6	0.13 (.882)
139N Awards	4.52 (1.76)	F7	4.61 (1.26)	F10	4.36 (1.86)	F3	4.54 (1.59)	F1	0.17 (.847)

Table 10. "Objective" validity ratings of 14 performance indicators

Variable	Total N = 116 Mean (SD)	Factorial Grouping	Administrs N = 16 Mean (SD)	Factorial Grouping	Teachers N = 66 Mean (SD)	Factorial Grouping	Researchers N = 24 Mean (SD)	Factorial Grouping	F (p)
I40A Experience	3.60 (2.30)	F1	3.36 (2.42)	F1	3.79 (2.23)	F1	3.22 (2.41)	F1	0.59 (.555)
I40B Grants	4.73 (1.71)	F2	5.73 (1.10)	F2	4.47 (1.82)	F2	4.65 (1.61)	F2	2.50 (.088)
I40C PR Conference Pres	5.15 (1.44)	F2	5.36 (1.12)	F2	5.15 (1.44)	F3	4.91 (1.47)	F3	0.43 (.655)
I40D NPR Conference	3.45 (1.75)	F1	3.08 (1.38)	F3	3.54 (1.66)	F4	3.30 (2.03)	F4	0.52 (.595)
I40E PR Journal	5.59 (1.33)	F2	5.82 (1.08)	F2	5.47 (1.40)	F5	5.61 (1.41)	F5	0.33 (.719)
I40F NPR Journal	3.48 (1.75)	F1	2.73 (1.35)	F3	3.58 (1.63)	F4	3.48 (2.06)	F4	1.13 (.327)
I40G Book Publication	5.36 (1.55)	F2	5.09 (1.04)	F4	5.30 (1.50)	F6	5.48 (1.83)	F6	0.24 (.784)
I40H Community Service	4.02 (1.81)	F3	4.18 (1.33)	F5	4.27 (1.82)	F7	3.48 (1.90)	F7	1.65 (.198)
I40I Students-supervised	5.79 (2.55)	F4	6.19 (3.17)	F6	5.75 (2.35)	F8	5.87 (2.72)	F8	0.19 (.828)
I40J Graduate Success	1.16 (0.37)	F5	1.06 (0.25)	F7	1.12 (0.33)	F9	1.29 (0.46)	F9	2.62 (.077)
I40K Courses-taught	4.54 (4.43)	F6	4.50 (3.20)	F2	4.06 (2.41)	F10	4.27 (2.75)	F10	0.19 (.823)
I40L Course Evaluation	1.07 (0.25)	F5	1.063 (0.25)	F7	1.06 (0.24)	F11	1.12 (0.34)	F10	0.54 (.587)
I40M Reputation	5.14 (2.42)	F6	5.06 (2.43)	F2	5.18 (2.32)	F12	5.08 (2.68)	F11	0.03 (.975)
I40N Awards	1.05 (0.22)	F7	1.12 (0.34)	F8	1.03 (0.17)	F13	1.04 (0.20)	F12	1.29 (.280)

Table 11. Differences in "subjective" and "objective" validity ratings of 14 performance indicators

Variable	Total N = 116 Mean(d) t(d)	Administrs N = 16 Mean(d) t(d)	Teachers N = 66 Mean(d) t(d)	Researchers N = 24 Mean(d) t(d)	F (p)
140A Experience	-0.23	-0.36	-0.15	-0.48	0.25
	-1.13	-0.40	-0.61	-1.75	(.780)
140B Grants	-0.24	-0.10	-0.47	0.26	1.69
	-1.48	-0.43	-1.98	0.92	(.190)
140C PR Conference	-0.35	0.10	-0.39	-0.48	0.56
	-2.36+	0.36	-2.00+	-1.39	(.571)
140D NPR Conference	-0.26	0.30	-0.25	-0.76	2.36
	-1.86	0.71	-1.58	-2.31+	(.101)
140E PR Journal	0.22	-0.20	0.19	0.56	1.20
	1.63	-0.80	1.08	1.62	(.306)
140F NPR Journal	0.00	1.10	-0.17	-0.18	5.10
	0.00	2.09	-1.15	-0.78	(.008)
140G Book Publication	0.24	0.70	0.15	0.22	0.46
	1.49	1.48	0.67	0.68	(.634)
140H Community Service	0.11	-0.18	0.03	0.09	0.11
	0.66	-0.28	0.16	0.37	(.891)
140I Students-supervised	-1.89	-2.50+	-1.68	-2.59	0.99
	-6.62***	-2.57	-4.42***	-4.75***	(.376)
140J Graduate Success	2.71	2.50	2.64	2.62	0.03
	15.51***	5.33**	12.08***	6.24***	(.970)
140K Courses-taught	-0.71	-0.33	-0.11	-1.57	2.07
	-1.58	-0.33	-0.35	-2.19+	(.132)
140L Course Evaluation	4.70	4.58	4.80	4.48	0.27
	26.48***	8.67***	19.65***	13.03***	(.764)
140M Reputation	-1.14	-0.83	-1.24	-1.50	0.18
	-3.75*	-0.83	-3.01++	-2.54+	(.836)
140N Awards	4.48	4.46	4.35	4.50	0.07
	25.99***	12.09***	18.23	14.13	(.929)

+ = $p \leq .05$, ++ = $p \leq .01$, * = $p \leq .001$, ** = $p \leq .0001$

Table 12. Overall ratings of structural factors that characterize a performance evaluation system

Variable		Total N = 116 Mean (SD) Factorial Grouping	Administrs N = 16 Mean (SD) Factorial Grouping	Teachers N = 66 Mean (SD) Factorial Grouping	Researchers N = 24 Mean (SD) Factorial Grouping	F (p)
P41	Purposes	5.79 (2.55)	6.19 (3.17)	5.74 (2.35)	5.87 (2.72)	0.19 (.828)
C42	Control	4.54 (4.43)	4.50 (3.20)	4.06 (2.41)	4.27 (2.75)	0.19 (.823)
I43	Procedures	5.14 (2.42)	5.06 (2.43)	5.18 (2.32)	5.08 (2.68)	0.03 (.975)
S44	Satisfaction	5.03 (2.34)	4.56 (2.87)	5.25 (2.21)	4.96 (1.97)	0.62 (.539)

P41 = A continuum that describes the balance of purposes of performance evaluation from "Informational" (rated 1) to "Evaluative" (rated 10). C42 = A continuum that describes the balance of control of performance evaluation from "Internal" (rated 1) to "External" (rated 10). I43 = A continuum that describes the balance of standards of performance evaluation from "Professional standards" (rated 1) to "Institutional standards" (rated 10). S44 = your own level of overall satisfaction with the existing process of performance evaluation at your own institution from "Very dissatisfied" (rated 1) to "Completely satisfied" (rated 10).

Figure 1. Locations of Some Performance Indicators in Three Dimensional Grid (Purpose, Control and Process): A Hypothetical Example

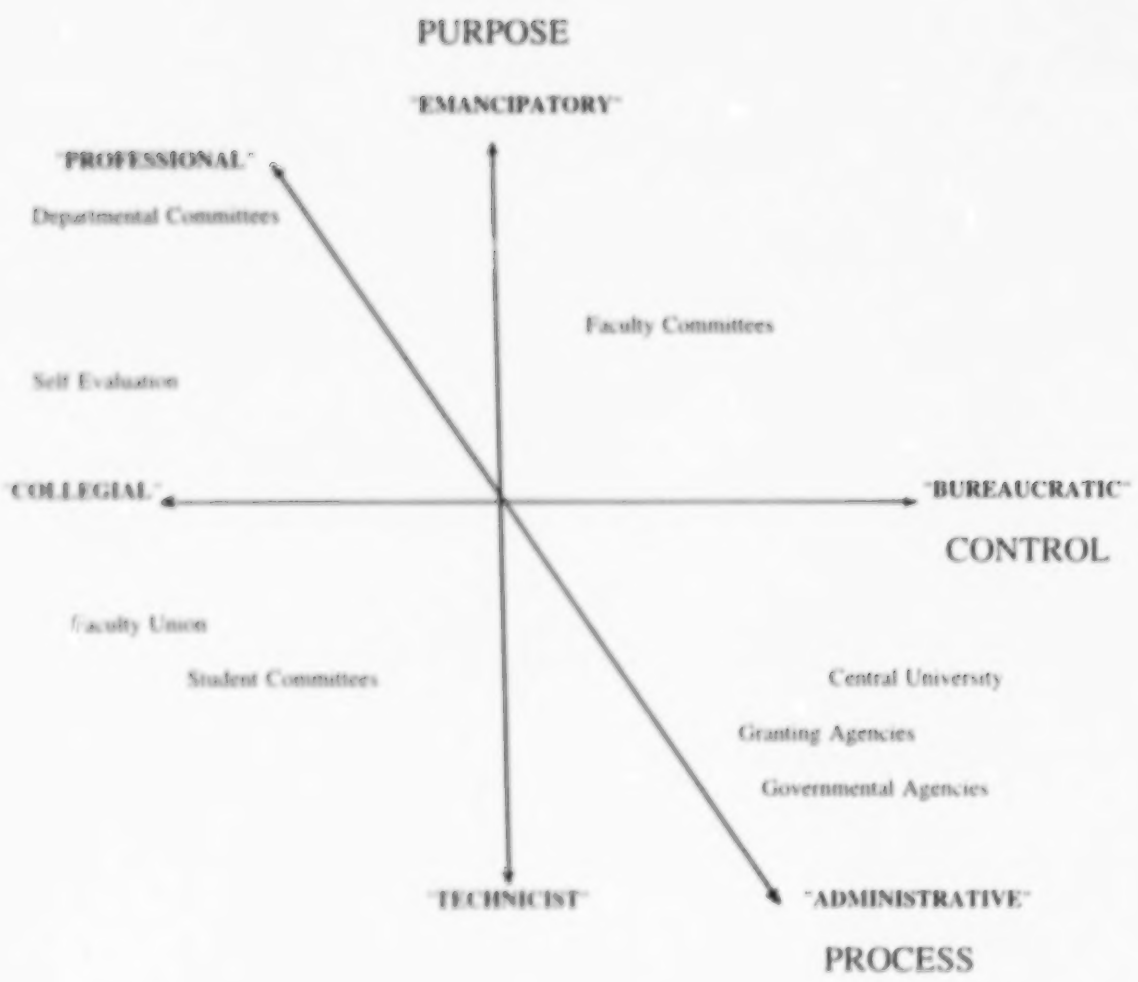
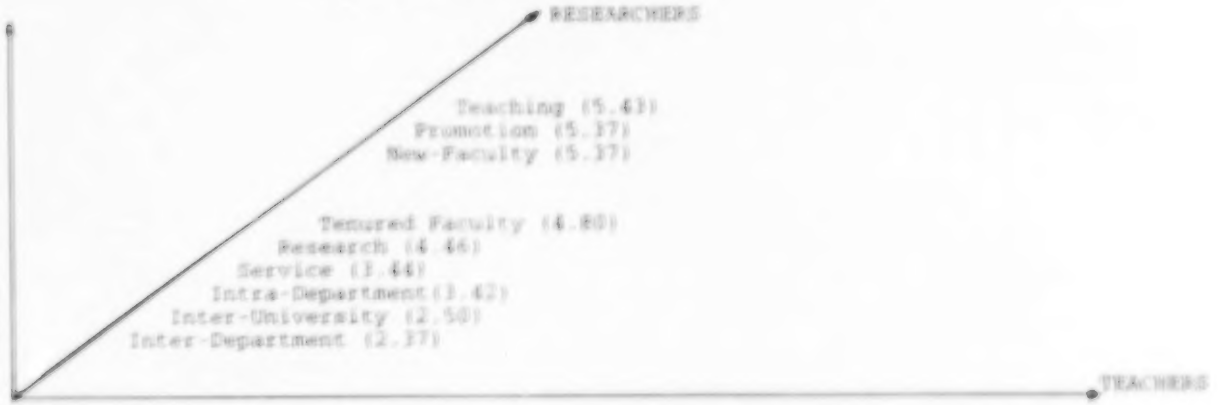

PURPOSE
"EMANCIPATORY"
"PROFESSIONAL"
Departmental Committees
Faculty Committees
Self Evaluation
"COLLEGIAL"
"BUREAUCRATIC"
CONTROL
Faculty Union
Student Committees
Central University
Granting Agencies
Governmental Agencies
"TECHNICIST"
"ADMINISTRATIVE"
PROCESS

Figure 2. Diagrams of Rank-ordered Means and Factorial Groupings in the Administrators,Teachers and Researchers Subsamples Corresponding to Tables 2 to 10

41

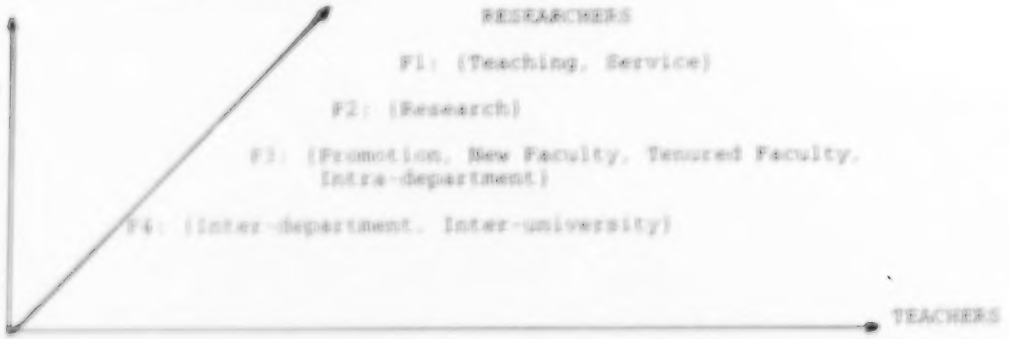
a. Relative Importance of Purposes Served by Performance Evaluation

Rank-ordered Means



Inter-department (2.48) Service (3.91) Tenured Faculty (4.83) Teaching (5.66)
Inter-university (2.17) Intra-Department (3.18) Research (4.23) New-Faculty (5.17)
Promotion (5.01)

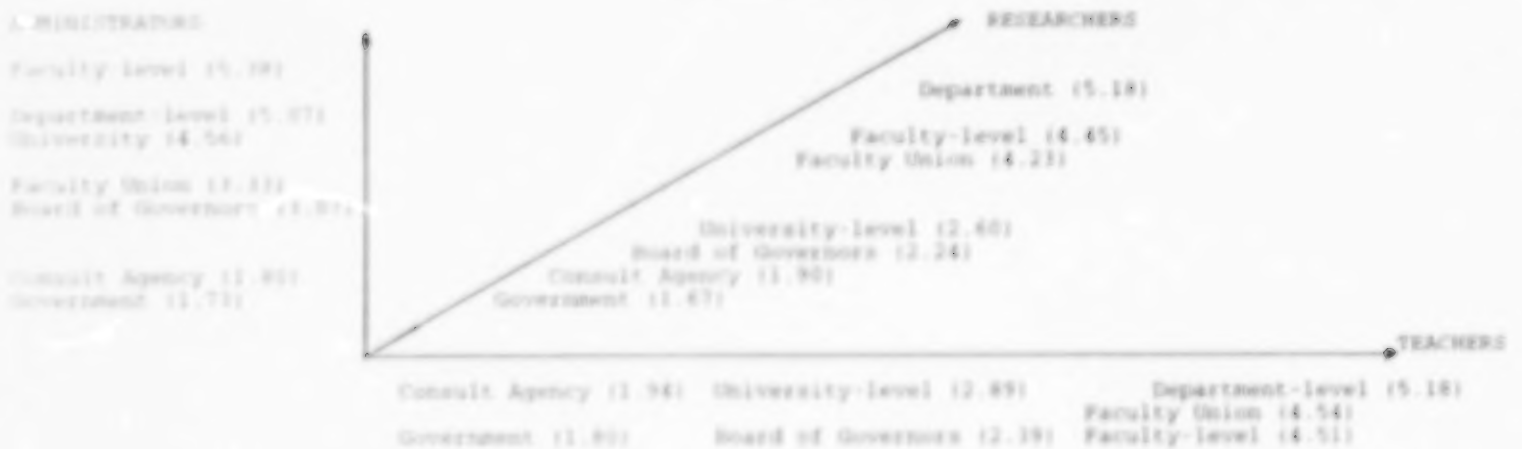
Factorial Groupings



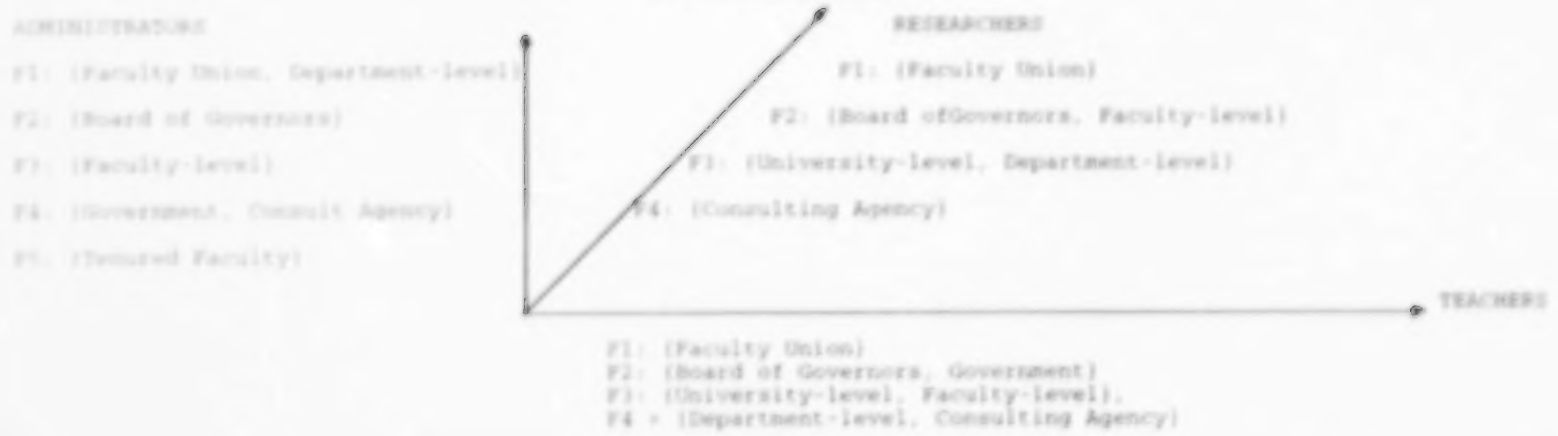
F1: (Teaching, Research, Service) F2: (Promotion, New Faculty)
F1: (Intra-department, Inter-department, Inter-university)

b. Relative Importance of Institutions that Determine Purposes of Performance Evaluation

Rank ordered Means

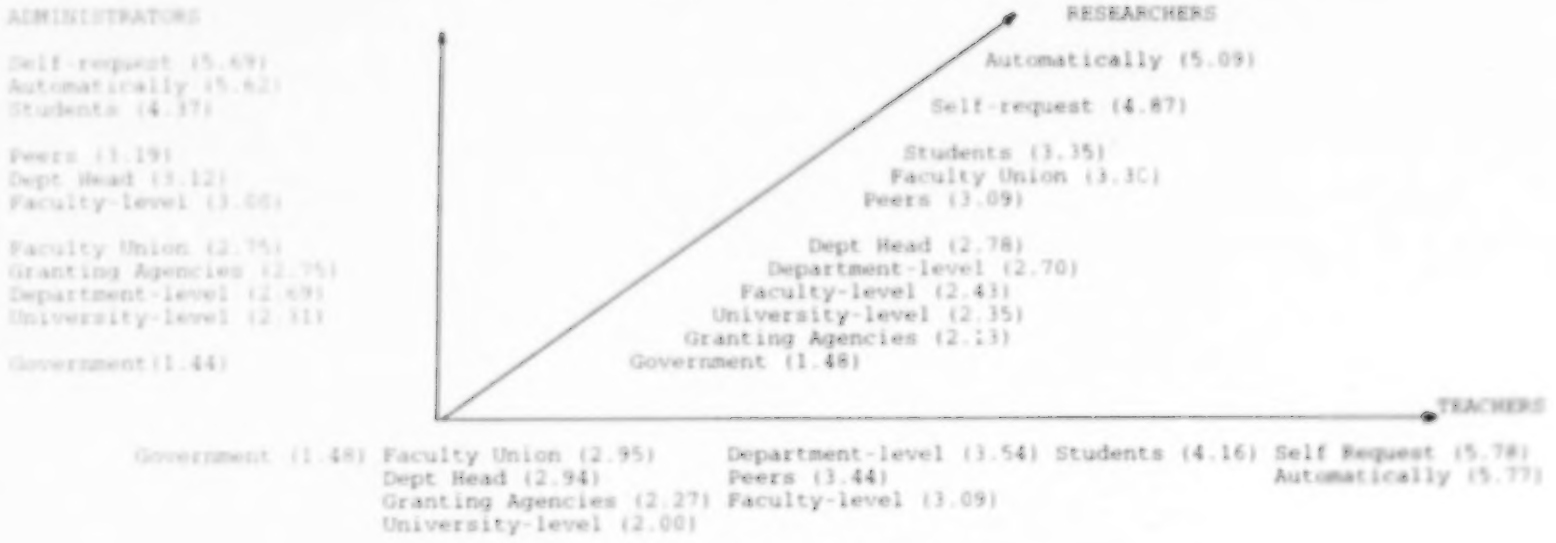


Factorial Groupings

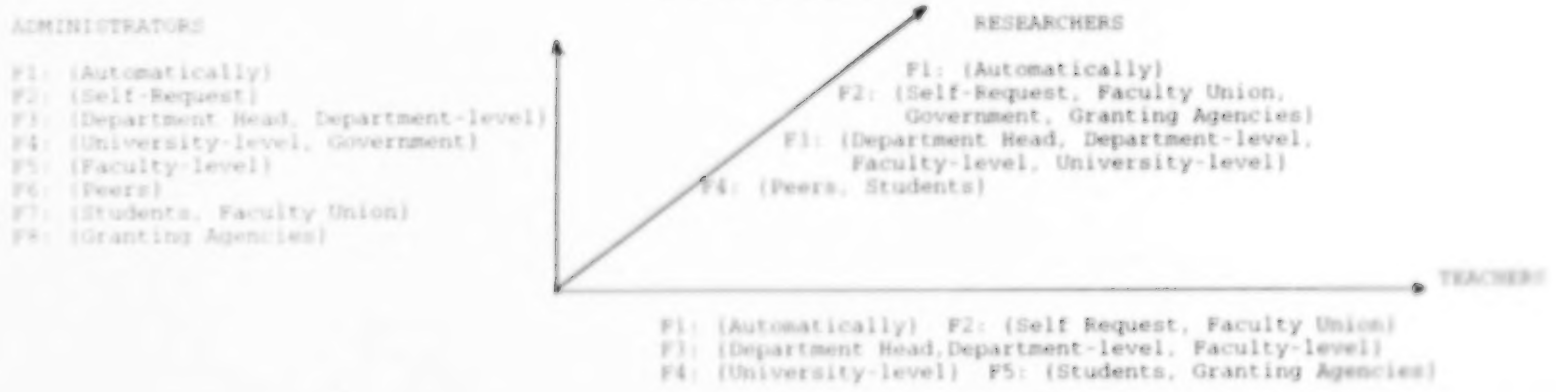


F1: (Faculty Union)
F2: (Board of Governors, Government)
F3: (University-level, Faculty-level),
F4 = (Department-level, Consulting Agency)

Rank-ordered Means

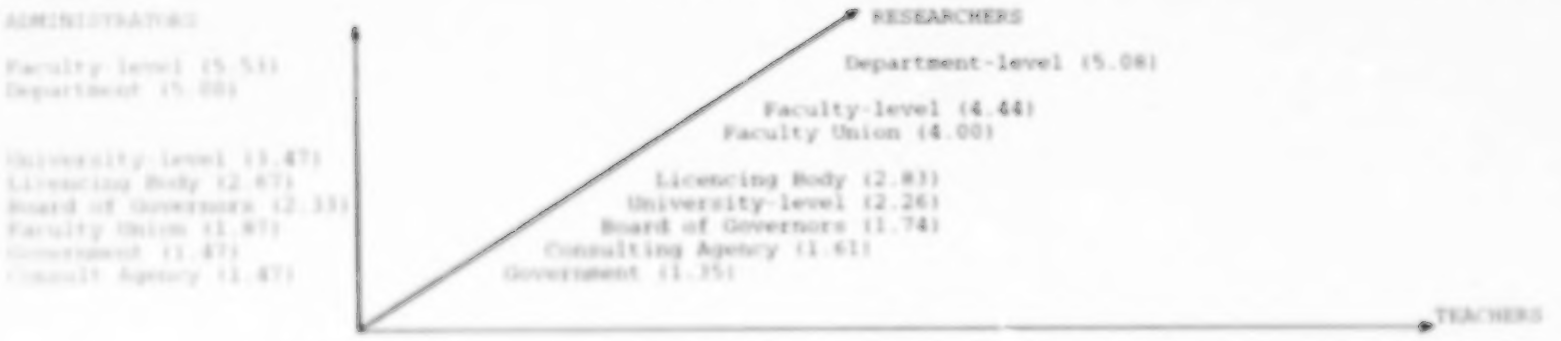


Factorial Groupings



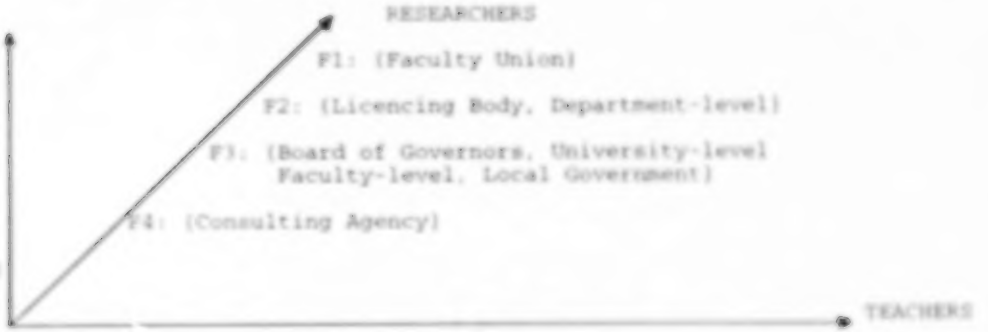
d. *Institutions Which Can Administer the Routine Process of Performance*

Rank-ordered Means



University-level (2.44) Licencing Agency (3.55) Faculty-level (4.59) Department-level (5.28)
Board of Governors (2.09) Faculty Union (1.52)
Consult Agency (1.97)
Government (1.26)

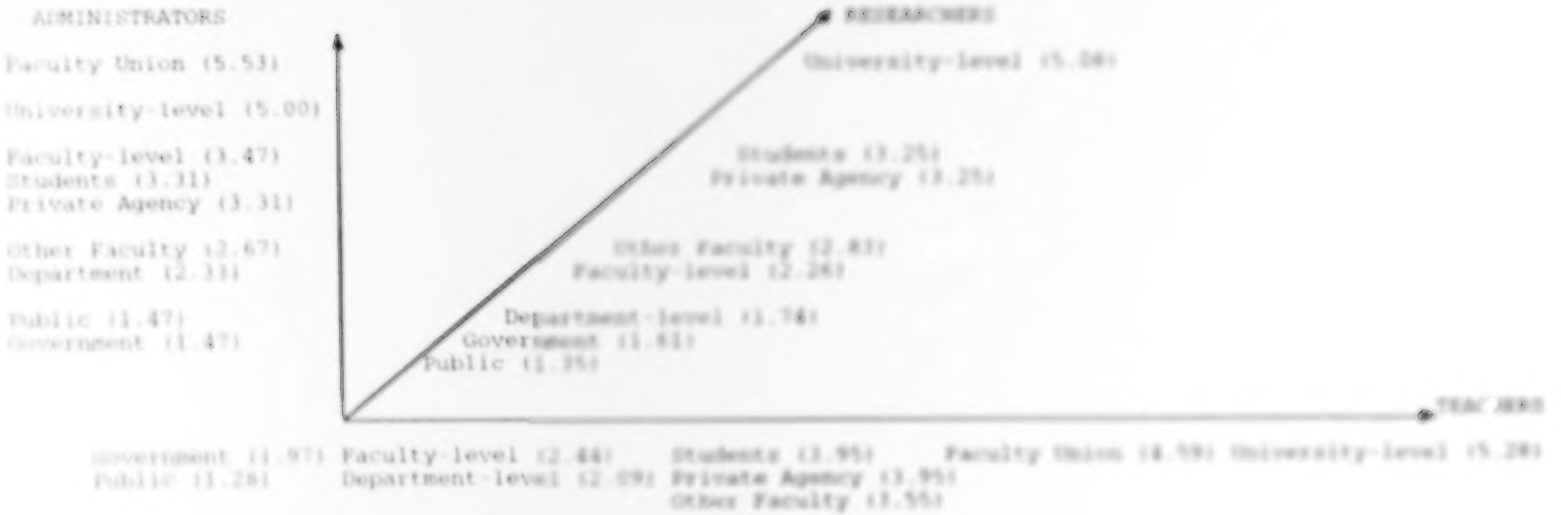
Factorial Groupings



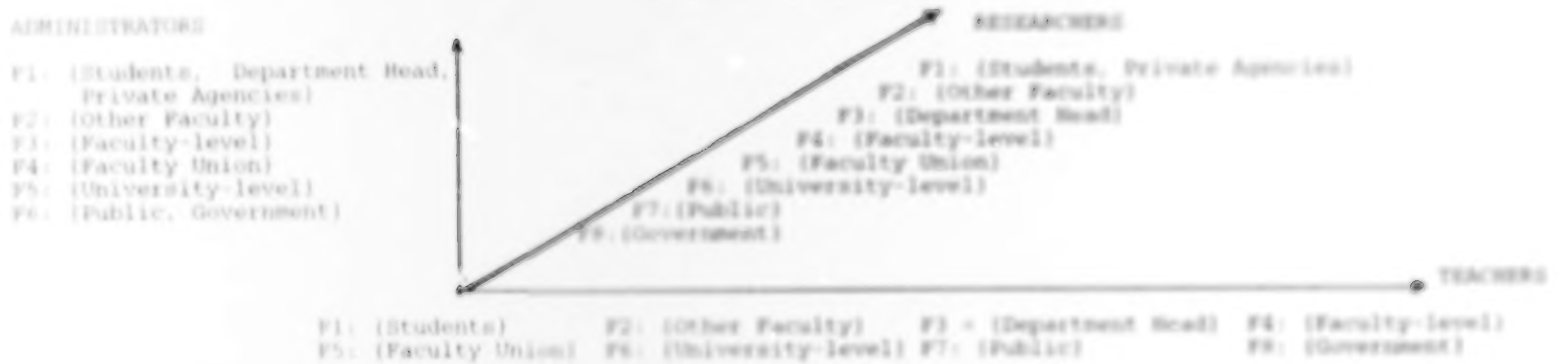
F1: (Faculty Union)
F2: (LicencingBody, Consulting Agency)
F3: (Board of Governors, University-level, Faculty-level)
F4: (Department-level, Local Government)

e. *Institutions Which Can Access Evaluation Information*

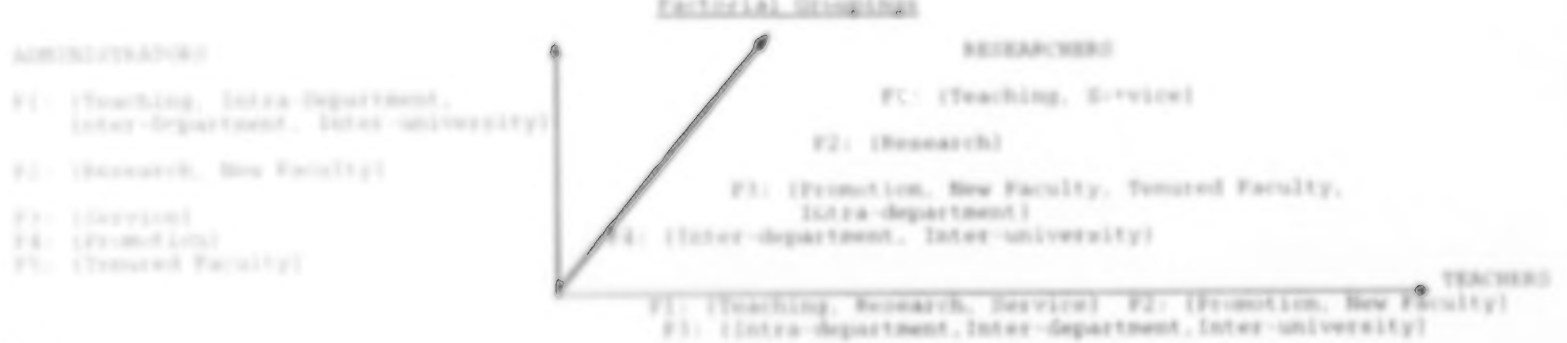
Rank-ordered Means



Factorial Groupings

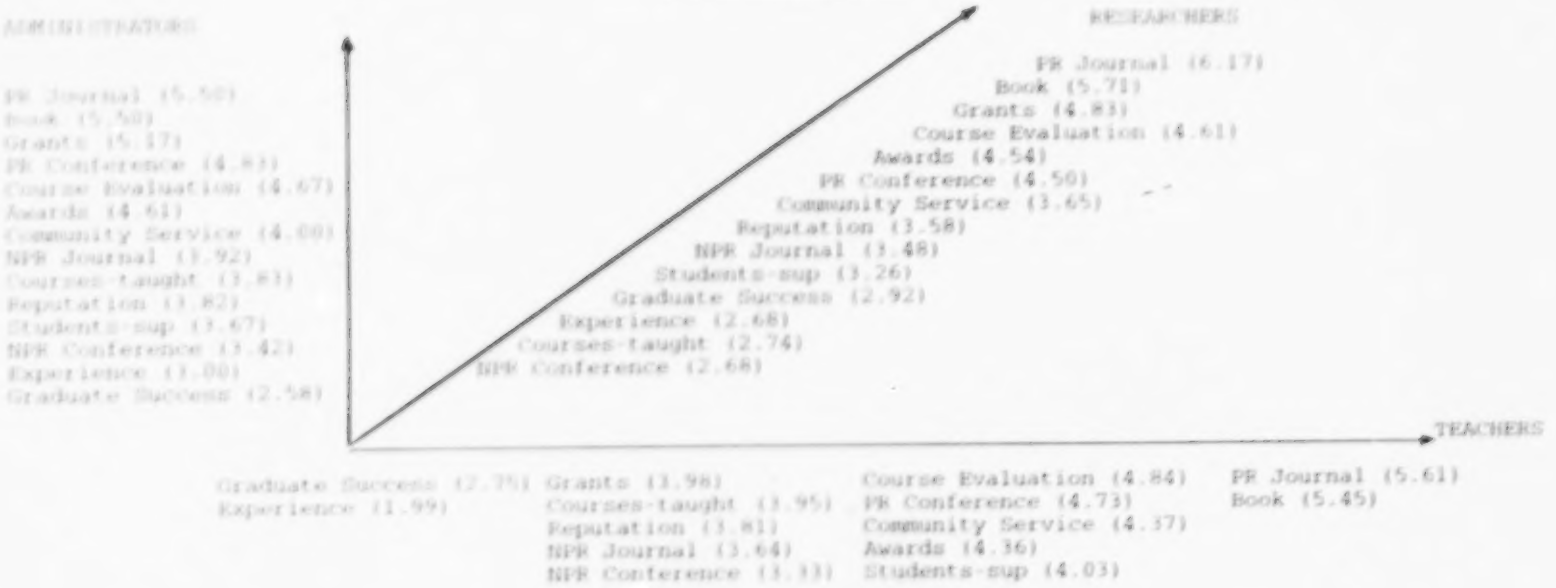


Rank-ordered Means
RESEARCHERS
PR Journal (.19)
Book (.15)
Course Evaluation (.12)
Grants (.11)
Experience (.08)
PR Conference (.04)
Awards (.06)
Students-sup (.05)
Community Service (.04)
Reputation (.03)
NPR Journals(.03)
Courses-taught(.03)
NPR Conference (.03)
Graduate Surveys (.02)
TEACHERS
Graduate (.02) NPR Conference (.04) Experience (.09) Course Evaluation (.15)
NPR Journal (.04) Community Service (.07) PR Journal (.13)
Reputation (.04) Grants (.07) Book (.11)
PR Conference (.07)
Awards (.06)
Courses-taught (.06)
Students-sup (.05)


Factorial Groupings
RESEARCHERS
F2: (Research)
F3: (Promotion, New Faculty, Tenured Faculty,
Intra-department)
F4: (Inter-department, Inter-university)
TEACHERS
F1: (Teaching, Research, Service) F2: (Promotion, New Faculty)
F3: (Intra-department,Inter-department,Inter-university)

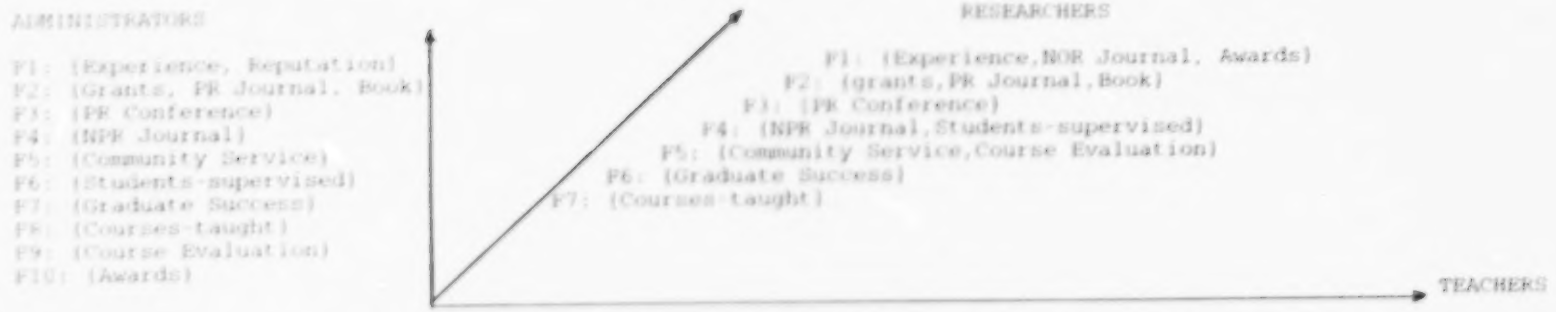
g. "Subjective" validity ratings of 14 Performance Indicators

Rank-ordered Means



Graduate Success (2.75) Grants (3.98) Course Evaluation (4.84) PR Journal (5.61)
Experience (1.99) Courses-taught (3.95) PR Conference (4.73) Book (5.45)
Reputation (3.81) Community Service (4.37)
NPR Journal (3.64) Awards (4.36)
NPR Conference (3.33) Students-sup (4.03)

Factorial Groupings

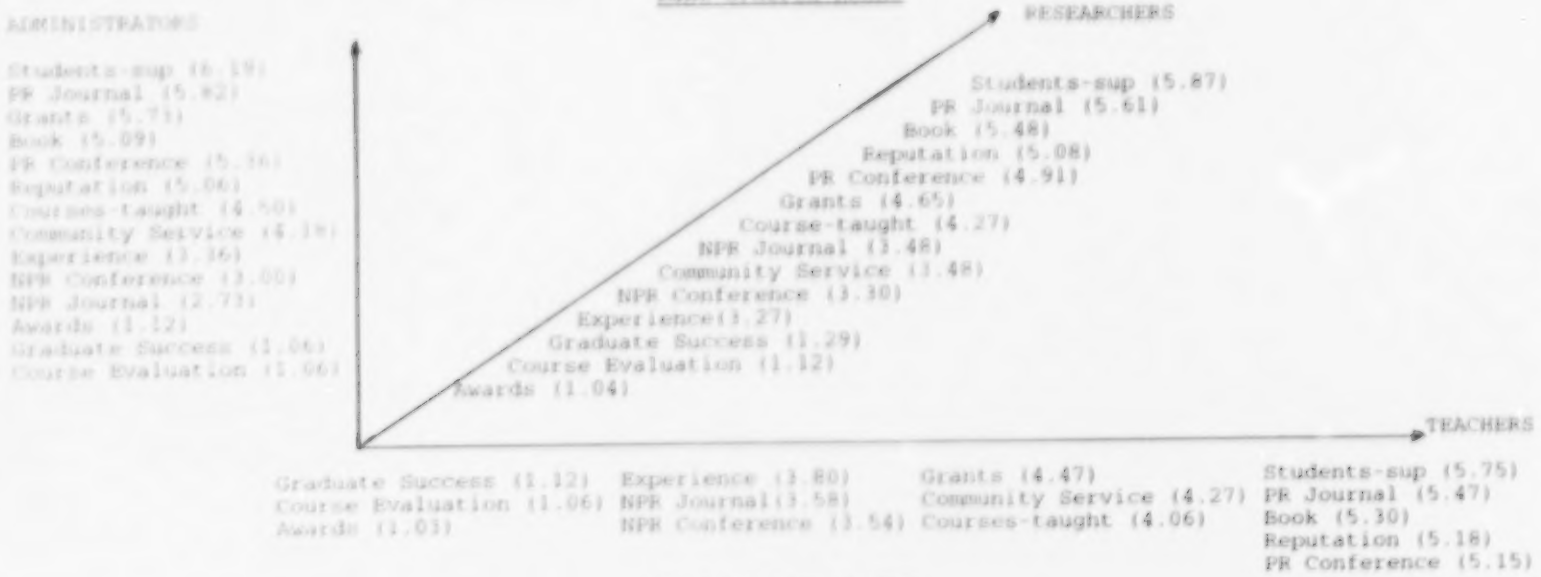


F1: (Experience) F2: (Grants, PR Juournal, Book) F3: (PR Conference)
F4: (NPR Conference) F5: (Community Service) F6:(Graduate Success, Reputation)

BEST COPY AVAILABLE

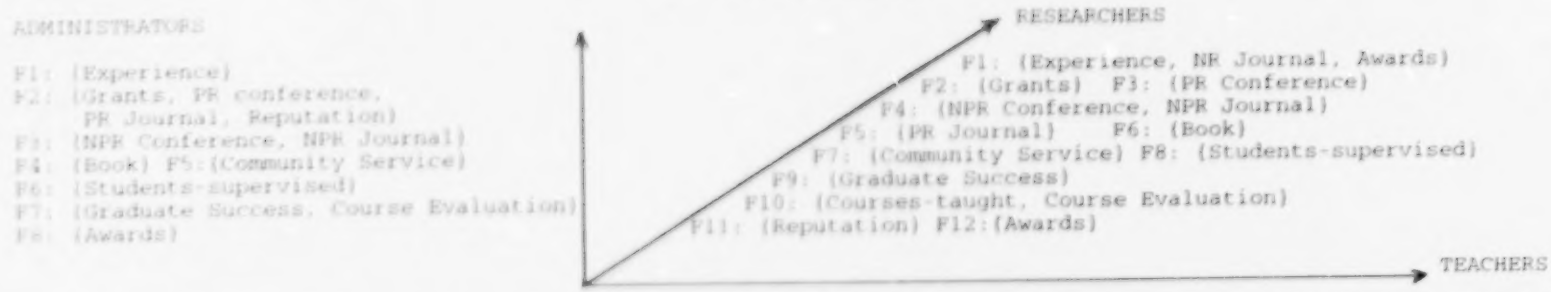
B. *"Objective" validity ratings of 14 Performance Indicators*

Rank-ordered Means



Graduate Success (1.12)	Experience (3.80)	Grants (4.47)	Students-sup (5.75)
Course Evaluation (1.06)	NPR Journal(3.58)	Community Service (4.27)	PR Journal (5.47)
Awards (1.03)	NPR Conference (3.54)	Courses-taught (4.06)	Book (5.30)
			Reputation (5.18)
			PR Conference (5.15)

Factorial Groupings



F1: (Experience)	F2: (Grants)	F3: (PR Conference)	F4: (NPR Conference, NPR Journal)	F5: (PR Journal)
F6: (Book)	F7: (Community Service)		F8: (Students-sup)	F9: (Graduate Success)
F10: (Course-taught)	F11: (Course Evaluation)		F12: (Reputation)	F13: (Awards)

BEST COPY AVAILABLE